

## **White Paper: The Digital Archives of Georgia**

### **Summary**

24 July 2006

**Purpose:** Manage the digital information of state agencies and, where appropriate, preserve the digital record for historical purposes

**Rationale:** Since 1732 the State of Georgia has carefully selected and preserved records of historical significance. Since 1918 these have been preserved in the state archives. Over the centuries, methods for selecting and preserving paper records have developed into a well-documented and cost-effective process. No such process yet exists for electronic records. Without conscious effort now, many of today's records that tell the story of Georgia will be lost to historians and the state's citizens.

**The Risk:** Electronic records can be read only if the proper hardware and software exists, and the pace with which technology changes make it unlikely that today's electronic records will be readable in even twenty or thirty years, let alone 200 years. Just 25 years ago, eight-inch floppy disks were still a common method of storing electronic files. Today, no computer in state government is capable of processing such a disk; and such a computer, if one were found, would no longer contain the software needed to read the files. Yet state agencies create records every day that must be available to citizens and historians for decades—and even centuries—into the future. Many records created in electronic form have long term legal, administrative, and historical value: Governor's correspondence files, e-mails, press releases, and Executive Orders; laws passed by the General Assembly (which shortly will be created and processed in electronic form); Department of Transportation maps and plans; even the photographs used by Tourism agencies to document and advertise the state—photos which have always provide a historical record of changing Georgia—are now created in electronic form and may well disappear within decades.

**The solution:** The Georgia Archives, in partnership with the Georgia Technology Authority, Pardons and Paroles, and other state agencies, has taken the first steps toward building a Digital Archive—one of the first in the nation. Using federal grant funds, the Archives drafted policies and procedures, piloted an entirely electronic Executive Clemency process, and installed the first server to harvest and preserve electronic documents. The development of a comprehensive Digital Archive will require the support of the Governor's Office, the General Assembly, and every state agency.

**The benefit:** The Digital Archive will manage electronic records so that each is kept as long as—and only as long as—its retention period requires. This will reduce storage costs, avoid potential litigation costs (by eliminating accidental loss of records and preventing the retention of records beyond their legal retention period), and ensure the survival of historical records in electronic form. In addition, the Digital Archive will build confidence in the state's ability to protect electronic records against accidental disclosure and loss, thereby encouraging the creation of more records in electronic form and reducing the state's dependence on paper records.

The pages which follow outline the lifecycle management of Georgia's electronic records in more detail. For more information, please contact the director of the Georgia Archives, David Carmicheal, at [dcarmicheal@sos.state.ga.us](mailto:dcarmicheal@sos.state.ga.us) or 678-364-3714.

# **White Paper: The Digital Archives of Georgia**

## **Lifecycle Management Overview**

24 July 2006

### **Lifecycle Management of Government Records in Georgia**

The Georgia Archives is mandated by statute to ensure the proper management and safeguarding of public records and to facilitate citizen and government access to these records. This mandate encompasses a wide range of responsibilities including preserving historical records, developing retention schedules, and ensuring the maintenance and security of all public records regardless of format. To accomplish its mission, the Georgia Archives has implemented a Lifecycle Management approach to government records and information.

Technology and the “e-government revolution” are having a substantial impact on the way government conducts its business and presents challenges for the capturing, preserving, managing, storing, and making accessible digital information. Significant amounts of critical digital data have already been lost. The primary purpose of the Digital Archives, as a component of the lifecycle management approach, is to preserve and provide access to records of enduring legal and historical significance. As government records are increasingly generated and stored in computer-based information systems, the state faces the challenge of managing and preserving this information for its own use in decision making and for the use of its citizenry.

### **Lifecycle Management of Information**

Why lifecycle management of information? Lifecycle management of information reinforces governance and managerial effectiveness. It improves internal audit and accountability procedures, facilitates rapid and accurate decision-making, sharpens an agency’s ability to respond to legal pressures, and furthers the ability of management to effectively assess work procedures. It also ensures that certain key records and information of historical, legal and fiscal significance are identified and preserved in a digital archives for further use by government.

Overall, less than three percent (3%) of the records and information maintained by state government would be preserved within the Digital Archives. The vast majority of information would be created, stored for a period of time, used to conduct the daily business of government, and then destroyed. The management of this information lifecycle serves to reduce the state’s exposure to liability and litigation, to ensure the effective use of the state’s technology investment, and to increase the efficiency of service to its citizens. In order to identify and ensure the preservation of this three percent, the Georgia Archives works with state agencies in better managing their information through

- Issuing retention schedules governing the maintenance of government records and identifying those few records of continuing historical, legal and fiscal value to the state
- Implementing enterprise content management technologies that enable the effective management of information and records
- Training agency staff and officials in the efficient management of information and the risks associated with maintaining information for too long a period of time, and
- Implementing a Digital Archives for the storage and preservation of key government records

### **Program Implementation**

Program implementation has focused in three areas:

- Development of standards and policy in support of data archiving, minimum metadata requirements, data transmission, and data sharing to ensure the creation of archives-ready and migration-ready records
- Establishment of a contract for the purchase of enterprise content management systems meeting the requirements of DoD 5015.2-STD
- Implementation of a Digital Archives system for the preservation of key government records

**Standards Development:** The issuance of standards and policy in support of data archiving, minimum metadata requirements, data transmission, and data sharing ensures the creation of archives-ready records that are more economical for the state to preserve over the long-term. However, this practice also saves the state funds in the short term by enabling the state to procure technology at lower costs due to volume purchasing (larger pool of agencies purchasing few products) and by facilitating the sharing of data between agencies (by eliminating the use of proprietary software and non-standard file formats).

**Enterprise Contract for ECM/RMA products:** The establishment of a contract for enterprise content management software meeting the requirements of DoD 5015.2-STD (a standard promulgated by the US Department of Defense) ensures that the records management components of retention schedules, secure destruction, and protection of confidential records are included in agency systems for the management of email, digital images, spreadsheets and other file formats.

**Digital Archives System Conceptual Overview:** The Digital Archives system will include various operational functions and processes for managing and accessing permanent digital objects and their associated metadata configured into three distinct storage components:

- Workspace component for temporary housing of objects during the records transfer process;
- Preservation component, aka the Digital Repository, for the long-term permanent storage and management of digital objects; and,
- Access component for temporary housing of records during use.

Unlying these three storage components are the administrative functions for managing the Digital Archives system. Each component of the Digital Archives is physically separate from the other, secured via firewalls and access restrictions.

**Workspace** storage encompasses the processes of Records Transfer – receiving, assuring quality, processing, importing, and certifying custody of digital objects. Digital objects will be transferred to the Digital Archives through File Transfer Protocol (FTP). They will be moved from the FTP server into an ‘inbox’ server for virus scanning and validation. A unique identifier consisting of a record id and an accession (shipment) number will be assigned at this point. Administrative and descriptive metadata will be added to and modified as the object moves through the various processes within each of the components of the Digital Archives System.

**Preservation** storage encompasses the processes of creating and managing metadata, retrieving data, maintaining data integrity, allowing for operations, storing digital objects, and managing storage media. Three versions of an object will be stored and monitored within the Preservation component, also known as the Digital Repository. The original object transferred by the agency will be encapsulated in XML and preserved along with a preservation copy stored as a non-proprietary file format. For example, if an agency transfers JPEG or PDF images to the Digital Archives System, the original JPEG or PDF will be preserved and a TIFF version created as the preservation copy. A third version for publication to the Digital Archives web page will also be created. This version may be redacted to conceal confidential information.

**Access** storage encompasses the processes needed to provide access to information – browsing and searching, indexing, responding to queries, displaying objects, ordering objects, registering users, and managing access. The public web interface is the primary architectural component of this function. Members of the public will interface with the Digital Archives System solely through the Access component. Only logged web addresses will be allowed to query the storage array for the retrieval of descriptive metadata and presentation copies. Presentation copies may be redacted to conceal confidential information.

**Conclusion:** Lifecycle management of information has substantial benefits to offer to state government agencies. It improves managerial ability to manage work processes, and to speed decision-making and response to citizen demands for service. It also enables the selection and preservation of key records and information for both the use of government officials in decision-making and the use of citizens exploring their state’s past. Working together the Georgia Archives and GTA have made significant progress but much remains to be accomplished.

- Work to establish a contract for the purchase of enterprise content management software has begun but remains to be completed.
- Additional standards in the areas of technology use, e-authentication and data archiving remain to be developed, and
- Full implementation of the Digital Archives system and the transfer of key records and information into the custody of the Archives must be completed. The Digital Archives currently remains in a pilot/prototype phase as we develop full requirements for the system.